Probabilistic Bisimulation Metrics

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Estonia Winter School March 2015

Outline

Process Equivalence is Fundamental

- Markov chains:
- Lumpability
- Labelled Markov processes: Bisimulation
- Markov decision processes: Bisimulation
- Labelled Concurrent Markov Chains with τ transitions: Weak Bisimulation

But...

- In the context of probability is exact equivalence reasonable?
- We say “no”. A small change in the probability distributions may result in bisimilar processes no longer being bisimilar though they may be very “close” in behaviour.
- Instead one should have a (pseudo)metric for probabilistic processes.

Bisimulation

- Let R be an equivalence relation. R is a bisimulation if: $s\ R\ t$ if $(\forall\ a)$:

$$(s \stackrel{a}{\to} P) \Rightarrow [t \stackrel{a}{\to} Q, P =_R Q]$$

$$(t \stackrel{a}{\to} Q) \Rightarrow [s \stackrel{a}{\to} P, P =_R Q]$$

- s, t are bisimilar if there is a bisimulation relating them.
- There is a maximum bisimulation relation.

Properties of Bisimulation

- Establishing equality of states: Coinduction. Establish a bisimulation R that relates states s, t.
- Distinguishing states: Simple logic is complete for bisimulation.

$$\phi ::= \mathtt{true} \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_{>q} \phi$$

- Bisimulation is sound for much richer logic pCTL*.
- Bisimulation is a congruence for usual process operators.

A metric-based approximate viewpoint

- Move from equality between processes to distances between processes (Jou and Smolka 1990).
- Formalize distance as a metric:

$$d(s,s) = 0, d(s,t) = d(t,s), d(s,u) \leq d(s,t) + d(t,u).$$

  Quantitative analogue of an equivalence relation.
- Quantitative measurement of the distinction between processes.

Summary of results

- Establishing closeness of states: Coinduction
- Distinguishing states: Real-valued modal logics
- Equational and logical views coincide: Metrics yield same distances as real-valued modal logics
- Compositional reasoning by *Non-Expansivity*. Process-combinators take nearby processes to nearby processes.

$$\frac{d(s_1, t_1) < \epsilon_1, \qquad d(s_2, t_2) < \epsilon_2}{d(s_1 \mid\mid s_2, t_1 \mid\mid t_2) < \epsilon_1 + \epsilon_2}$$

- Results work for Markov chains, Labelled Markov processes, Markov decision processes and Labelled Concurrent Markov chains with τ-transitions.

Criteria on Metrics

- Soundness:
$$d(s,t) = 0 \Leftrightarrow \ s,t \text{ are bisimilar}$$
- Stability of distance under temporal evolution:“Nearby states stay close *forever*.”
- Metrics should be computable (efficiently?).

Bisimulation Recalled

Let R be an equivalence relation. R is a bisimulation if: $s\ R\ t$ if:

$$(s \longrightarrow P) \Rightarrow [t \longrightarrow Q, P =_R Q]$$

$$(t \longrightarrow Q) \Rightarrow [s \longrightarrow P, P =_R Q]$$

where $P =_R Q$ if

$$(\forall R - \text{closed } E)\ P(E) = Q(E)$$

A putative definition of a metric-bisimulation

- m is a metric-bisimulation if: $m(s,t) < \epsilon \Rightarrow$:

$$s \longrightarrow P \Rightarrow\ t \longrightarrow Q, \quad m(P,Q) < \epsilon$$

$$t \longrightarrow Q \Rightarrow\ s \longrightarrow P, \ \ m(P,Q) < \epsilon$$

- Problem: what is $m(P,Q)$? — Type mismatch!!
- Need a way to lift distances from states to a distances on distributions of states.

A detour: Kantorovich-Wasserstein metric

- Metrics on probability measures on metric spaces.
- $\mathcal{M}$: 1-bounded pseudometrics on states.
- $$d(\mu, \nu) = \sup_f |\int f d\mu - \int f d\nu|, f \text{ 1-Lipschitz}$$
- Arises in the solution of an LP problem: *transshipment*.

An LP version for Finite-State Spaces

When state space is finite: Let P, Q be probability distributions. Then:

$$m(P,Q) = \max \sum_i (P(s_i) - Q(s_i))a_i$$

subject to:

$$\begin{aligned}&\forall i. 0 \leq a_i \leq 1\\ &\forall i,j.\ a_i - a_j \leq m(s_i, s_j).\end{aligned}$$

The Dual Form

- Dual form from Worrell and van Breugel:
- $$\min \sum_{i,j} l_{ij} m(s_i, s_j) + \sum_i x_i + \sum_j y_j$$

  subject to:

  $$\begin{array}{l} \forall i.\, \sum_j l_{ij} + x_i = P(s_i) \\ \forall j.\, \sum_i l_{ij} + y_j = Q(s_j) \\ \forall i,j.\ l_{ij}, x_i, y_j \geq 0. \end{array}$$
- We prove many equations by using the primal form to show one direction and the dual to show the other.

Example 1

- $m(P,P) = 0$.
- In dual, match each state with itself, $l_{ij} = \delta_{ij} P(s_i), x_i = y_j = 0$. So:

$$\sum_{i,j} l_{ij} m(s_i, s_j) + \sum_i x_i + \sum_j y_j$$

  becomes 0.
- This clearly cannot be lowered further so this is the min.

Example 2

- Let $m(s,t) = r < 1$. Let $\delta_s(\delta_t)$ be the probability measure concentrated at $s(t)$. Then,

$$m(\delta_s, \delta_t) = r$$

- Upper bound from dual: Choose $l_{st} = 1$ all other $l_{ij} = 0$. Then

$$\sum_{ij} l_{ij} m(s_i, s_j) = m(s,t) = r.$$

- Lower bound from primal: Choose $a_s = 0, a_t = r$, all others to match the constraints. Then

$$\sum_i (\delta_t(s_i) - \delta_s(s_i)) a_i = r.$$

The Importance of Example 2

We can *isometrically* embed the original space in the metric space of distributions.

Example 3 - I

- Let $P(s) = r, P(t) = 0$ if $s \neq t$. Let $Q(s) = r', Q(t) = 0$ if $s \neq t$.
- Then $m(P,Q) = |r - r'|$.
- Assume that $r \geq r'$.
  Lower bound from primal: yielded by $\forall i. a_i = 1$,

$$\sum_i (P(s_i) - Q(s_i)) a_i = P(s) - Q(s) = r - r'.$$

Example 3 - II

Upper bound from dual: $l_{ss} = r'$ and $x_s = r - r'$, all others 0

$$\sum_{i,j} l_{ij} m(s_i, s_j) + \sum_i x_i + \sum_j y_j = x_s = r - r'.$$

and the constraints are satisfied:

$$\sum_j l_{sj} + x_s = l_{ss} + x_s = r$$

$$\sum_i l_{is} + y_s = l_{ss} = r'.$$

Return from Detour

Summary of detour: Given a metric on states in a metric space, can lift to a metric on probability distributions on states.

Metric "Bisimulation"

- m is a metric-bisimulation if: $m(s,t) < \epsilon \Rightarrow$:

$$s \longrightarrow P \Rightarrow \ t \longrightarrow Q, \quad m(P,Q) < \epsilon$$

$$t \longrightarrow Q \Rightarrow \ s \longrightarrow P, \ \ m(P,Q) < \epsilon$$

- The required canonical metric on processes is the least such: ie. the distances are the least possible.
- Thm: *Canonical least metric exists*. Usual fixed-point theory arguments.

Metrics: some details

- $\mathcal{M}$: 1-bounded pseudometrics on states with ordering

$$m_1 \preceq m_2 \text{ if } (\forall s,t)\ [m_1(s,t) \geq m_2(s,t)]$$

- $(\mathcal{M}, \preceq)$ is a complete lattice.
-

$$\begin{aligned}
\bot(s,t) &= \begin{cases} 0 \text{ if } s = t \\ 1 \text{ otherwise} \end{cases} \\
\top(s,t) &= 0, (\forall s,t) \\
(\sqcap\{m_i\}(s,t) &= \sup_i m_i(s,t)
\end{aligned}$$

Maximum fixed point definition

- Let $m \in \mathcal{M}$. $F(m)(s,t) < \epsilon$ if:

$$s \longrightarrow P \Rightarrow \; t \longrightarrow Q, \quad m(P,Q) < \epsilon$$

$$t \longrightarrow Q \Rightarrow \; s \longrightarrow P, \;\; m(P,Q) < \epsilon$$

- $F(m)(s,t)$ can be given by an explicit expression.
- F is monotone on $\mathcal{M}$, and metric-bisimulation is the greatest fixed point of F.
- The closure ordinal of F is ω.

A Key Tool: Splitting

Let P and Q be probability distributions on a set of states. Let P_1 and P_2 be such that: $P = P_1 + P_2$. Then, there exist Q_1, Q_2, such that $Q_1 + Q_2 = Q$ and

$$m(P, Q) = m(P_1, Q_1) + m(P_2, Q_2).$$

The proof uses the duality theory of LP.

What about Continuous-State Systems?

- Develop a real-valued “modal logic” based on the analogy:

Program Logic	Probabilistic Logic
State s	Distribution μ
Formula ϕ	Random Variable f
Satisfaction $s \models \phi$	$\int f \mathrm{d}\mu$

- Define a metric based on how closely the random variables agree.
- We did this *before* the LP based techniques became available.

Real-valued Modal Logic

- $$f ::= \mathbf{1} \mid \max(f,f) \mid h \circ f \mid \langle a \rangle f$$

- $$\begin{array}{lcll} \mathbf{1}(s) & = & 1 & \text{True} \\ \max(f_1,f_2)(s) & = & \max(f_1(s),f_2(s)) & \text{Conjunction} \\ h \circ f(s) & = & h(f(s)) & \text{Lipschitz} \\ \langle a \rangle f(s) & = & \gamma \int_{s' \in S} f(s') \tau_a(s, \mathrm{d}s') & a\text{-transition} \end{array}$$

  where h 1-Lipschitz : $[0,1] \to [0,1]$ and $\gamma \in (0,1]$.
- $d(s,t) = \sup_f |f(s) - f(t)|$
- Thm: d coincides with the canonical metric-bisimulation.

Finitary syntax for Real-valued modal logic

$$
\begin{array}{lcll}
\mathbf{1}(s) & = & 1 & \text{True} \\
\max(f_1,f_2)(s) & = & \max(f_1(s),f_2(s)) & \text{Conjunction} \\
(1-f)(s) & = & 1-f(s) & \text{Negation} \\
\lfloor f_q(s) \rfloor & = & \begin{cases} q \ , & f(s) \geq q \\ f(s) \ , & f(s) < q \end{cases} & \text{Cutoffs} \\
\langle a \rangle f(s) & = & \gamma \int_{s' \in S} f(s') \tau_a(s, \mathrm{d}s') & a\text{-transition}
\end{array}
$$

q is a rational.

The role of γ

- γ discounts the value of future steps.
- $\gamma < 1$ and $\gamma = 1$ yield very different topologies
- The approximants defined last week converge in the metric $\gamma < 1$.
- The $\gamma < 1$ metric yields a topology in which many more sequences converge.
- For $\gamma < 1$ there is an LP-based strongly-polynomial (in the number of constraints, and the number of bits of precision required) algorithm to compute the metric.
- For $\gamma = 1$ the existence of an algorithm to compute the metric has been discovered by van Breugel et al.

Conclusions

- For a CSP-like process algebra (without hiding) the process combinators are all contractive.
- We can show that if one perturbs the probabilities slightly the resulting process is close to the unperturbed one.
- We have an asymptotic version of the metric.
- We can extend the LP-based theory to continuous state spaces using the theory of infinite dimensional LP.